

A Probabilistic Algorithm for Computing Data-Discriminants of Likelihood Equations

Jose Israel Rodriguez and Xiaoxian Tang

June 13, 2016

Abstract

An algebraic approach to the maximum likelihood estimation problem is to solve a very structured parameterized polynomial system called likelihood equations that have finitely many complex (real or non-real) solutions. The only solutions that are statistically meaningful are the real solutions with positive coordinates. In order to classify the parameters (data) according to the number of real/positive solutions, we study how to efficiently compute the discriminants, say data-discriminants (DD), of the likelihood equations. We develop a probabilistic algorithm with three different strategies for computing DDs. Our implemented probabilistic algorithm based on `Maple` and `FGb` is more efficient than our previous version [27] and is also more efficient than the standard elimination for larger benchmarks. By applying `RAGlib` to a DD we compute, we give the real root classification of 3 by 3 symmetric matrix model.

Key Words: Maximum likelihood estimation, Likelihood equation, Discriminant

1 Introduction

In this work, we address the statistics problem of *maximum likelihood estimation* (MLE). A statistical model is a subset of \mathbb{R}^{n+1} whose points have nonnegative coordinates summing to one; these points represent probability distributions. We will be interested in statistical models that are defined by zero sets of polynomials restricted to the positive orthant. The study of such models is a central part of algebraic statistics.

Given a data set and statistical model, we ask, “Which point in the model best describes the data?” One way to answer this question is by maximum likelihood estimation. By determining the point in the model that maximizes the likelihood function, one finds the probability distribution that best describes the data.

For the models we consider, the data is discrete; more specifically, the data is a list of integers representing counts. The likelihood function we consider on the model is given by the monomial

$$\ell_u(p) := p_0^{u_0} p_1^{u_1} \cdots p_n^{u_n},$$

where $u = (u_0, \dots, u_n)$ is our data.

The *maximum likelihood estimate* is the point in the statistical model that maximizes the likelihood function. To determine this estimate, one can solve the likelihood equations to determine critical points of the likelihood function. This set of critical points contains all local maxima. So when the maximum likelihood estimate is a local maximum, it will be one of these critical points. When the maximum likelihood estimate is not a local maximum, then it is in the boundary (in the Euclidean topology) or the singular locus of the model. In those cases, one restricts to those loci’s Zariski closure and performs an analysis in this restriction. Iterating this procedure, one solves the global optimization problem of MLE.

The number of complex solutions to the likelihood equations (for general data) is a topological invariant called the *maximum likelihood degree*. For discrete models, the maximum likelihood degree

was shown to be an Euler-characteristic in the smooth case [22] and recently shown to be a weighted sum of Euler characteristics in the singular case [1]. In this paper, we will only consider discrete data. For a recent survey for the ML degree, one can refer to [23].

The ML degree measures the algebraic complexity of maximum likelihood estimation by counting the number of complex solutions (real and nonreal solutions) to the likelihood equations. In statistics only real solutions with positive coordinates are relevant, and to analyze the real root classification problem, one studies a discriminant of the system. Discriminants are a large topic studied in algebraic geometry, and they are important because they are used in combination with Ehresmann’s Theorem (Theorem 2) for real root classification. In principle, discriminants can be computed with Gröbner bases to determine elimination ideals [7, Chapter 3] or with geometric resolutions [13]. In this paper we give a probabilistic interpolation algorithm to determine discriminants of the likelihood equations. These interpolation methods can be applied to other parameterized system of equations, for example determining the Euclidean distance degree, which was defined in [8].

Our main contribution is a probabilistic algorithm (Algorithm 2) that is more efficient for larger-sized models than the standard algorithm (Algorithm 1). Tables 1–2 show our contribution is nontrivial with great performance increases in particular for Model 4. Based on the discriminants we have computed by the probabilistic algorithm, we give the real root classification for 3 by 3 symmetric matrix model. The main differences between this paper and our ISSAC’15 paper [27] are listed below.

- We have reimplemented [27, Algorithms 1–2] based on `Maple` and `FGb`. Comparing [27, Tables 1–2] and Tables 1–2 in Section 4, we see the new implementation performs much better. A typical example is Model 4. [27] states the old implementation of Algorithm 2 (Strategy 2) takes 30 days to compute the discriminant while our new implementation takes less than 30 minutes.
- In Section 3, we add Section 3.2.3 in which we discuss the motivations and effectiveness of different strategies. We add Strategy 3 which improves the efficiency of sampling over Strategy 2 in some cases. We also add Lemma 3 in Section 3.2.1, which guarantees the correctness of Strategy 3.
- In Section 4, we present the new timings in Tables 1–2 based on our new implementation. We try more examples such as Dense Models 3–6 and Models 5–9. We add Table 3 which compares Strategy 2 and Strategy 3.

Now we motivate our study with an illustrating example.

Example 1 (Illustrative Example). *Suppose we have a weighted four-sided die such that the probability p_i of observing side i ($i = 0, 1, 2, 3$) of the die satisfies the constraint $p_0 + 2p_1 + 3p_2 - 4p_3 = 0$. We toss the die 1000 times and record a 4-dimensional data vector (u_0, u_1, u_2, u_3) , where u_i is the number of times we observe the side i . We want to determine the probability distribution $(p_0, p_1, p_2, p_3) \in \mathbb{R}_{>0}^4$ that best explains the data subject to the constraint by maximum likelihood estimation:*

Maximize the likelihood function $p_0^{u_0} p_1^{u_1} p_2^{u_2} p_3^{u_3}$ subject to

$$p_0 + 2p_1 + 3p_2 - 4p_3 = 0, p_0 + p_1 + p_2 + p_3 = 1,$$

$$p_0 > 0, p_1 > 0, p_2 > 0, \text{ and } p_3 > 0.$$

For some statistical models, the MLE problem can be solved by well known local hill climbing algorithms such as the expectation–maximization algorithm. However, the local method can fail if there is more than one local maximum. One way to remedy this, is by solving the system of likelihood equations [19, 3]:

$$\begin{aligned} F_0 &= p_0\lambda_1 + p_0\lambda_2 - u_0 & F_3 &= p_3\lambda_1 - 4p_3\lambda_2 - u_3 \\ F_1 &= p_1\lambda_1 + 2p_1\lambda_2 - u_1 & F_4 &= p_0 + 2p_1 + 3p_2 - 4p_3 \\ F_2 &= p_2\lambda_1 + 3p_2\lambda_2 - u_2 & F_5 &= p_0 + p_1 + p_2 + p_3 - 1 \end{aligned}$$

where λ_1 and λ_2 are newly introduced indeterminates (Lagrange multipliers) for formulating the likelihood equations. More specifically, for given (u_0, u_1, u_2, u_3) , if (p_0, p_1, p_2, p_3) is a critical point

of the likelihood function, then there exist numbers λ_1 and λ_2 such that $(p_0, p_1, p_2, p_3, \lambda_1, \lambda_2)$ is a solution of the polynomial system.

For general data $u = (u_0, u_1, u_2, u_3)$, the likelihood equations have 3 complex solutions. However, only solutions with positive coordinates p_i are statistically meaningful. A solution with all positive p_i coordinates is said to be a positive solution. Real root classification (RRC) and positive root classification (PRC) are important problems:

For which u , the polynomial system has 1, 2 and 3 real/positive solutions?

According to the theory of computational (real) algebraic geometry [33, 25], the number of (real/positive) solutions only changes when the data u_i goes across some “special” values (see Theorem 2). The set of “special” u_i is a (projective) variety [25, Lemma 4] in (3 dimensional complex projective space) 4-dimensional complex space.

The number of real/positive solutions is uniform over each open connected component determined by the variety. In other words, the “special” u_i plays the similar role as the discriminant for univariate polynomials. The first step of RRC is calculating the “special” u_i , leading to the discriminant problem:

How to effectively compute the “special” u_i ?

Geometrically, the “special” u_i is a projection of a variety. So it can be computed by elimination. For instance, with the command `fgb_gbasis_elim` in FGB [10], we compute that the “special” u_i in Example 2 form a hypersurface defined by a homogenous polynomial in u_0, u_1, u_2, u_3 . However, for many MLE problems, the elimination computation is too expensive as shown in Tables 1–2 in Section 4.

We remark that our work can be viewed as following the numerous efforts in applying computational algebraic geometry to tackle MLE and critical points problems [19, 3, 2, 20, 31, 16, 12, 17, 23, 18, 26].

The paper is organized as follows. Section 2 proves the discriminant variety of likelihood equation is projective and gives the definition of data-discriminant. Section 3 presents the elimination algorithm and probabilistic algorithm (with three strategies). Section 4 shows the experimental results comparing the two algorithms and different strategies. Section 5 discusses the future work and gives the real root classification of 3 by 3 symmetric matrix model.

2 Definitions and preliminaries

In this section, we discuss how to define “data-discriminant”. We assume the reader is familiar with elimination theory [7, Chapter 3].

Notation 1. Let \mathbb{P} denote the projective closure of the complex numbers \mathbb{C} . For homogeneous polynomials g_1, \dots, g_s in $\mathbb{Q}[p_0, \dots, p_n]$, $\mathcal{V}(g_1, \dots, g_s)$ denotes the projective variety in \mathbb{P}^n defined by g_1, \dots, g_s . Let Δ_n denote the n -dimensional probability simplex $\{(p_0, \dots, p_n) \in \mathbb{R}^{n+1} | p_0 > 0, \dots, p_n > 0, p_0 + \dots + p_n = 1\}$.

Definition 1. (Algebraic Statistical Model and Model Invariant) The set \mathcal{M} is said to be an algebraic statistical model if $\mathcal{M} = \mathcal{V}(g_1, \dots, g_s) \cap \Delta_n$ where g_1, \dots, g_s define an irreducible generically reduced projective variety. Each g_i is said to be a model invariant of \mathcal{M} . If the codimension of $\mathcal{V}(g_1, \dots, g_s)$ is k , we say $\{h_1, \dots, h_k\}$ is a set of general model invariants for \mathcal{M} whenever the variety $\mathcal{V}(h_1, \dots, h_k)$ has $\mathcal{V}(g_1, \dots, g_s)$ as an irreducible component.

For a given algebraic statistical model, there are several different ways to formulate the likelihood equations [19]. In this section, we introduce the Lagrange likelihood equations and define the data-discriminant for this formulation. One can similarly define data-discriminants for other formulations of likelihood equations.

Notation 2. For any f_1, \dots, f_m in the polynomial ring $\mathbb{Q}[x_1, \dots, x_k]$, $\mathcal{V}_a(f_1, \dots, f_m)$ denotes the affine variety in \mathbb{C}^k defined by f_1, \dots, f_m and $\langle f_1, \dots, f_m \rangle$ denotes the ideal generated by f_1, \dots, f_m . For an ideal I in $\mathbb{Q}[x_1, \dots, x_k]$, $\mathcal{V}_a(I)$ denotes the affine variety defined by I .

Definition 2. (Lagrange Likelihood Equations and Correspondence) Given an algebraic statistical model X of codimension k and a set of general model invariants $\{h_1, \dots, h_k\}$, the system of polynomials below is said to be the Lagrange likelihood equations of X when set to zero:

$$\begin{aligned} F_0 &:= p_0(\lambda_1 + \frac{\partial h_1}{\partial p_0} \lambda_2 + \dots + \frac{\partial h_k}{\partial p_0} \lambda_{k+1}) - u_0 \\ &\vdots \\ F_n &:= p_n(\lambda_1 + \frac{\partial h_1}{\partial p_n} \lambda_2 + \dots + \frac{\partial h_k}{\partial p_n} \lambda_{k+1}) - u_n \\ F_{n+1} &:= g_1(p_0, \dots, p_n) \\ &\vdots \\ F_{n+s} &:= g_s(p_0, \dots, p_n) \\ F_{n+s+1} &:= p_0 + \dots + p_n - 1 \end{aligned}$$

where g_1, \dots, g_s are the model invariants of X and $u_0, \dots, u_n, p_0, \dots, p_n, \lambda_1, \dots, \lambda_{k+1}$ are indeterminates (also denoted by $\mathbf{u}, \mathbf{p}, \Lambda$). More specifically, $p_0, \dots, p_n, \lambda_1, \dots, \lambda_{k+1}$ are unknowns and u_0, \dots, u_n are parameters.

$\mathcal{V}_a(F_0, \dots, F_{n+s+1})$, namely the set

$$\{(\mathbf{u}, \mathbf{p}, \Lambda) \in \mathbb{C}^{n+1} \times \mathbb{C}^{n+1} \times \mathbb{C}^{k+1} \mid F_0 = 0, \dots, F_{n+s+1} = 0\},$$

is said to be the Lagrange likelihood correspondence of X and denoted by \mathcal{L}_X .

Notation 3. Let π denote the canonical projection from the ambient space of the Lagrange likelihood correspondence to the \mathbb{C}^{n+1} associated to the \mathbf{u} indeterminants $\pi: \mathbb{C}^{n+1} \times \mathbb{C}^{n+s+2} \rightarrow \mathbb{C}^{n+1}$. For any set S in \mathbb{C}^{n+1} , $\mathcal{I}(S)$ denotes the ideal

$$\{D \in \mathbb{Q}[\mathbf{u}] \mid D(a_0, \dots, a_n) = 0, \forall (a_0, \dots, a_n) \in S\}.$$

\overline{S} denotes the affine closure of S in \mathbb{C}^{n+1} , namely $\mathcal{V}_a(\mathcal{I}(S))$.

Given an algebraic statistical model \mathcal{M} and a data vector $\mathbf{u} \in \mathbb{R}_{>0}^n$, the *maximum likelihood estimation* problem is to **maximize the likelihood function** $p_0^{u_0} \dots p_n^{u_n}$ **subject to** \mathcal{M} . The MLE problem can be solved by computing $\pi^{-1}(\mathbf{u}) \cap \mathcal{L}_X$. More specifically, if \mathbf{p} is a regular point of $\mathcal{V}(g_1, \dots, g_s)$, then \mathbf{p} is a critical point of the likelihood function if and only if there exists $\Lambda \in \mathbb{C}^{k+1}$ such that $(\mathbf{u}, \mathbf{p}, \Lambda) \in \mathcal{L}_X$. Theorem 1 states that for a general data vector \mathbf{u} , $\pi^{-1}(\mathbf{u}) \cap \mathcal{L}_X$ is a finite set and the cardinality of $\pi^{-1}(\mathbf{u}) \cap \mathcal{L}_X$ is constant over a dense Zariski open set, which inspires the definition of ML degree. For details, see [19].

Theorem 1. [19] For an algebraic statistical model X , there exists an affine variety $V \subset \mathbb{C}^{n+1}$ and a non-negative integer N such that for any $\mathbf{u} \in \mathbb{C}^{n+1} \setminus V$,

$$\#\pi^{-1}(\mathbf{u}) \cap \mathcal{L}_X = N.$$

Definition 3. [19](**ML Degree**) For an algebraic statistical model X , the non-negative integer N stated in Theorem 1 is said to be the ML degree of X .

Definition 4. For an algebraic statistical model X with a set of general model invariants $\{h_1, \dots, h_k\}$, suppose F_0, \dots, F_{n+s+1} are defined as in Definition 2. Then, we have the following:

- \mathcal{L}_{X_∞} denotes the set of non-properness of π , i.e., the set of the $u \in \pi(\mathcal{L}_X)$ such that there does not exist a compact neighborhood U of u where $\pi^{-1}(U) \cap \mathcal{L}_X$ is compact;
- \mathcal{L}_{X_p} denotes $\pi(\mathcal{L}_X \cap \mathcal{V}_a(\Pi_{k=0}^n p_k))$; and

• $\mathcal{L}_{X,J}$ denotes $\overline{\pi(\mathcal{L}_X \cap \mathcal{V}_a(J))}$ where J denotes the determinant below

$$\det \begin{bmatrix} \frac{\partial F_0}{\partial p_0} & \cdots & \frac{\partial F_0}{\partial p_n} & \frac{\partial F_0}{\partial \lambda_1} & \cdots & \frac{\partial F_0}{\partial \lambda_{k+1}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F_n}{\partial p_0} & \cdots & \frac{\partial F_n}{\partial p_n} & \frac{\partial F_n}{\partial \lambda_1} & \cdots & \frac{\partial F_n}{\partial \lambda_{k+1}} \\ 1 & \cdots & 1 & 0 & \cdots & 0 \\ \frac{\partial h_1}{\partial p_0} & \cdots & \frac{\partial h_1}{\partial p_n} & \frac{\partial h_1}{\partial \lambda_1} & \cdots & \frac{\partial h_1}{\partial \lambda_{k+1}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_k}{\partial p_0} & \cdots & \frac{\partial h_k}{\partial p_n} & \frac{\partial h_k}{\partial \lambda_1} & \cdots & \frac{\partial h_k}{\partial \lambda_{k+1}} \end{bmatrix}_{(n+k+2) \times (n+k+2)}.$$

The geometric meaning of \mathcal{L}_{X_p} and $\mathcal{L}_{X,J}$ are as follows. The first, \mathcal{L}_{X_p} , is the projection of the intersection of the Lagrange likelihood correspondence with the coordinate hyperplanes. The second, $\mathcal{L}_{X,J}$, is the projection of the intersection of the Lagrange likelihood correspondence with the hypersurface defined by J . Geometrically, $\mathcal{L}_{X,J}$ is the closure of the union of the projection of the singular locus of \mathcal{L}_X and the set of critical values of the restriction of π to the regular locus of \mathcal{L}_X [25, Definition 2].

The Lagrange likelihood equations define an affine variety. As we continuously deform the parameters u_i , coordinates of a solution can tend to infinity. Geometrically, \mathcal{L}_{X_∞} is the set of the data \mathbf{u} such that the Lagrange likelihood equations have some solution (\mathbf{p}, Λ) at infinity; this is the closure of the set of “non-properness” as defined in the page 1, [24] and page 3, [29]. It is known that the set of non-properness of π is closed and can be computed by Gröbner bases (see Lemma 2 and Theorem 2 in [25]).

The ML degree captures the geometry of the likelihood equations over the complex numbers. However, statistically meaningful solutions occur over real numbers. Below, Theorem 2 states that \mathcal{L}_{X_∞} , $\mathcal{L}_{X,J}$ and \mathcal{L}_{X_p} define open connected components such that the number of real/positive solutions is uniform over each open connected component. Theorem 2 is a corollary of *Ehresmann’s theorem* for which there exists semi-algebraic statements since 1992 [6].

Theorem 2. *For an algebraic statistical model X . If \mathcal{C} is an open connected component of*

$$\mathbb{R}^{n+1} \setminus (\mathcal{L}_{X_\infty} \cup \mathcal{L}_{X,J}),$$

then over $\mathbf{u} \in \mathcal{C}$, the following is constant:

$$\#\pi^{-1}(\mathbf{u}) \cap \mathcal{L}_X \cap \mathbb{R}^{n+s+2}.$$

Moreover, if \mathcal{C} is an open connected component of

$$\mathbb{R}^{n+1} \setminus (\mathcal{L}_{X_\infty} \cup \mathcal{L}_{X,J} \cup \mathcal{L}_{X_p}),$$

then over $\mathbf{u} \in \mathcal{C}$, the following is constant:

$$\#\pi^{-1}(\mathbf{u}) \cap \mathcal{L}_X \cap (\mathbb{R}_{>0}^{n+1} \times \mathbb{R}^{s+1}).$$

Before we give the definition of data-discriminant, we study the structures of \mathcal{L}_{X_p} , $\mathcal{L}_{X,J}$ and \mathcal{L}_{X_∞} . Proposition 1 shows that the structure of the likelihood equations forces \mathcal{L}_{X_p} to be contained in the union of coordinate hyperplanes defined by $\prod_{k=0}^n u_k$. Proposition 2 shows that the structure of the likelihood equations forces $\mathcal{L}_{X,J} \setminus \{\mathbf{0}\}$ to be a projective variety. Similar to the proof of Proposition 2, we can also show that the structure of the likelihood equations forces $\mathcal{L}_{X_\infty} \setminus \{\mathbf{0}\}$ to be a projective variety.

Proposition 1. *For any algebraic statistical model X ,*

$$\mathcal{L}_{X_p} \subset \mathcal{V}_a(\prod_{k=0}^n u_k).$$

Proof. By Definition 2, for any k ($0 \leq k \leq n$),

$$u_k = p_k(\lambda_1 + \frac{\partial g_1}{\partial p_1} \lambda_2 + \cdots + \frac{\partial g_s}{\partial p_1} \lambda_{s+1}) - F_k.$$

Hence,

$$u_k \in \langle F_k, p_k \rangle \cap \mathbb{Q}[u_k] \subset \langle F_0, \dots, F_{n+s+1}, p_k \rangle \cap \mathbb{C}[\mathbf{u}]$$

So

$$\mathcal{V}_a(\langle F_0, \dots, F_{n+s+1}, p_k \rangle \cap \mathbb{C}[\mathbf{u}]) \subset \mathcal{V}_a(u_k)$$

By the Closure Theorem [7],

$$\mathcal{V}_a(\langle F_0, \dots, F_{n+s+1}, p_k \rangle \cap \mathbb{C}[\mathbf{u}]) = \overline{\pi(\mathcal{L}_X \cap \mathcal{V}_a(p_k))}$$

Therefore,

$$\begin{aligned} \mathcal{L}_{X_p} &= \overline{\pi(\mathcal{L}_X \cap \mathcal{V}_a(\Pi_{k=0}^n p_k))} \\ &= \overline{\pi(\mathcal{L}_X \cap \cup_{k=0}^n \mathcal{V}_a(p_k))} \\ &= \cup_{k=0}^n \overline{\pi(\mathcal{L}_X \cap \mathcal{V}_a(p_k))} \\ &\subset \cup_{k=0}^n \mathcal{V}_a(u_k) \\ &= \mathcal{V}_a(\Pi_{k=0}^n u_k). \square \end{aligned}$$

Remark 1. Generally, $\mathcal{L}_{X_p} \neq \mathcal{V}_a(\Pi_{k=0}^n u_k)$. For example, suppose the algebraic statistical model is $\mathcal{V}_a(p_0 - p_1) \cap \Delta_1$. Then $\mathcal{L}_{X_p} = \emptyset \neq \mathcal{V}_a(u_0 u_1)$.

Notation 4. \mathcal{D}_{X_p} denotes the product $\Pi_{k=0}^n u_k$.

Proposition 2. For an algebraic statistical model X , we have $\mathcal{L}_{X_J} \setminus \{\mathbf{0}\}$ is a projective variety in \mathbb{P}^n , where $\mathbf{0}$ is the zero vector $(0, \dots, 0)$ in \mathbb{C}^{n+1} .

Proof. By the formulation of the Lagrange likelihood equations, we can prove that $\mathcal{I}(\pi(\mathcal{L}_X \cap \mathcal{V}_a(J)))$ is a homogeneous ideal by the two basic facts below, which can be proved by Definition 2 and basic algebraic geometry arguments.

F1. For every \mathbf{u} in $\pi(\mathcal{L}_X \cap \mathcal{V}_a(J))$, each scalar multiple $\alpha \mathbf{u}$ is also in $\pi(\mathcal{L}_X \cap \mathcal{V}_a(J))$.

F2. For any $S \subset \mathbb{C}^{n+1}$, if for any $\mathbf{u} \in S$ and for any scalar $\alpha \in \mathbb{C}$, $\alpha \mathbf{u} \in S$, then $\mathcal{I}(S)$ is a homogeneous ideal in $\mathbb{Q}[\mathbf{u}]$.

That means the ideal $\mathcal{I}(\pi(\mathcal{L}_X \cap \mathcal{V}_a(J)))$ is generated by finitely many homogeneous polynomials D_1, \dots, D_m . Therefore, $\mathcal{L}_{X_J} = \mathcal{V}_a(\mathcal{I}(\pi(\mathcal{L}_X \cap \mathcal{V}_a(J)))) = \mathcal{V}_a(D_1, \dots, D_m)$. So $\mathcal{L}_{X_J} \setminus \{\mathbf{0}\} = \mathcal{V}(D_1, \dots, D_m) \subset \mathbb{P}^n$. \square

Notation 5. For an algebraic statistical model X , we define the notation \mathcal{D}_{X_J} according to the codimension of $\mathcal{L}_{X_J} \setminus \{\mathbf{0}\}$ in \mathbb{P}^n .

- If the codimension is 1, then assume $\mathcal{V}(D_1), \dots, \mathcal{V}(D_K)$ are the codimension 1 irreducible components in the minimal irreducible decomposition of $\mathcal{L}_{X_J} \setminus \{\mathbf{0}\}$ in \mathbb{P}^n and $\langle D_1 \rangle, \dots, \langle D_K \rangle$ are radical. \mathcal{D}_{X_J} denotes the homogeneous polynomial $\Pi_{j=1}^K D_j$.

- If the codimension is greater than 1, then our convention is to take $\mathcal{D}_{X_J} = 1$.

Similarly, we use the notation \mathcal{D}_{X_∞} to denote the projective variety $\mathcal{L}_{X_\infty} \setminus \{\mathbf{0}\}$. Now we define the “data-discriminant” of Lagrange likelihood equations.

Definition 5. (Data-Discriminant) For a given algebraic statistics model X , the homogeneous polynomial $\mathcal{D}_{X_\infty} \cdot \mathcal{D}_{X_J} \cdot \mathcal{D}_{X_p}$ is said to be the data-discriminant (DD) of Lagrange likelihood equations of X and denoted by \mathcal{D}_X .

Remark 2. Note that DD can be viewed as a generalization of the “discriminant” for univariate polynomials. So it is interesting to compare DD with border polynomial (BP) [33] and discriminant variety (DV) [25]. DV and BP are defined for general parametric polynomial systems. DD is defined for the likelihood equations but can be generalized to generic zero-dimensional systems. Generally, for any square and generic zero-dimensional system, $\mathcal{V}_a(DD) \subset DV \subset \mathcal{V}_a(BP)$. Note that due to the special structure of likelihood equations, DD is a homogenous polynomial despite being an affine system of equations. However, generally, DV is not a projective variety and BP is not homogenous.

Example 2 (Linear Model). The algebraic statistic model for the four sided die story in Section 1 is given by

$$\mathcal{M} = \mathcal{V}(p_0 + 2p_1 + 3p_2 - 4p_3) \cap \Delta_3.$$

The Langrange likelihood equations are the $F_0 = 0, \dots, F_5 = 0$ shown in Example 1. The Langrange likelihood correspondence is $\mathcal{L}_X = \mathcal{V}_a(F_0, \dots, F_5) \subset \mathbb{C}^{10}$. If we choose generic $(u_0, u_1, u_2, u_3) \in \mathbb{C}^4$, $\pi^{-1}(u_0, u_1, u_2, u_3) \cap \mathcal{L}_X = 3$, namely the ML degree is 3. The data-discriminant is the product of $\mathcal{D}_{X\infty}$, \mathcal{D}_{Xp} and \mathcal{D}_{XJ} , where

$$\begin{aligned} \mathcal{D}_{X\infty} &= u_0 + u_1 + u_2 + u_3, \quad \mathcal{D}_{Xp} = u_0 u_1 u_2 u_3, \quad \text{and} \\ \mathcal{D}_{XJ} &= 441u_0^4 + 4998u_0^3u_1 + 20041u_0^2u_1^2 + 33320u_0u_1^3 + 19600u_1^4 - 756u_0^3u_2 + 20034u_0^2u_1u_2 + 83370u_0u_1^2u_2 + 79800u_1^3u_2 - \\ &5346u_0^2u_2^2 + 55890u_0u_1u_2^2 + 119025u_1^2u_2^2 + 4860u_0u_2^3 + 76950u_1u_2^3 + 18225u_2^4 - 1596u_0^3u_3 - 11116u_0^2u_1u_3 - 17808u_0u_1^2u_3 + \\ &4480u_1^3u_3 + 7452u_0^2u_2u_3 - 7752u_0u_1u_2u_3 + 49680u_1^2u_2u_3 - 17172u_0u_2^2u_3 + 71460u_1u_2^2u_3 + 27540u_2^3u_3 + 2116u_0^2u_3^2 + 6624u_0u_1u_3^2 - \\ &4224u_1^2u_3^2 - 9528u_0u_2u_3^2 + 15264u_1u_2u_3^2 \\ &+ 14724u_2^2u_3^2 - 1216u_0u_3^3 - 512u_1u_3^3 + 3264u_2u_3^3 + 256u_3^4. \end{aligned}$$

By applying the well known partial cylindrical algebraic decomposition (PCAD) [5] method to the data-discriminant above, we get that for any $(u_0, u_1, u_2, u_3) \in \mathbb{R}_{>0}^4$,

- if $\mathcal{D}_{XJ}(u_0, u_1, u_2, u_3) > 0$, then the system of likelihood equations has 3 distinct real solutions and 1 of them is positive;
- if $\mathcal{D}_{XJ}(u_0, u_1, u_2, u_3) < 0$, then the system of likelihood equations has exactly 1 real solution and it is positive.

The answer above can be verified by the `RealRootClassification` [33, 4] command in `Maple 2015`. In this example, the $\mathcal{D}_{X\infty}$ does not effect the number of real/positive solutions since it is always positive when each u_i is positive. However, generally, $\mathcal{D}_{X\infty}$ plays an important role in real root classification. Also remark that the real root classification is equivalent to the positive root classification for this example but it is not true generally (see the example discussed in Section 5).

3 Algorithm

In this section, we discuss how to compute the discriminant \mathcal{D}_X . We assume that X is the closure of a given statistical model, F_0, \dots, F_{n+s+1} are defined as in Definition 2, and J is defined as in Definition 4. We rename F_0, \dots, F_{n+s+1} as F_0, \dots, F_m . We also rename $p_0, \dots, p_n, \lambda_1, \dots, \lambda_s$ as y_0, \dots, y_m . Subsection 3.1 presents the standard algorithm for reference and Subsection 3.2 presents the probabilistic algorithm.

3.1 Standard Algorithm

Considering the data-discriminant as a projection drives a natural algorithm to compute it. This is the standard elimination algorithm in symbolic computation:

- we compute the \mathcal{L}_{XJ} by *elimination* and then get \mathcal{D}_{XJ} by the *radical equidimensional decomposition* [25, Definition 3]. The algorithm is formally described in the Algorithm 1;
- we compute $\mathcal{L}_{X\infty}$ by the Algorithm `PROPERNESSDEFECTS` presented in [25] and then get $\mathcal{D}_{X\infty}$ by the radical equidimensional decomposition. We omit the formal description of the algorithm.

Practically, Algorithm 1 may not terminate in a reasonable time before a computer reaches its memory limit. For instance, when using `FGb` for our Gröbner Bases computations the memory limit is reached due to the large size of the intermediate computational results. Since the algorithms for

Algorithm 1: DX-J

input : F_0, \dots, F_m, J
output: \mathcal{D}_{XJ}
1 $\mathcal{G}_{\mathbf{u}} \leftarrow$ a set of generators of the elimination ideal $\langle F_0, \dots, F_m, J \rangle \cap \mathbb{Q}[\mathbf{u}]$
2 $\mathcal{D}_{XJ} \leftarrow$ the codimension 1 component of the equidimensional radical decomposition of $\langle \mathcal{G}_{\mathbf{u}} \rangle$
3 **return** \mathcal{D}_{XJ}

Gröbner Bases in **FGb** are based on row echelon form computations [11, Section 2], we record the sizes of matrices generated by **FGb** for Models 4–9, see Table 4 in Section 4. This motivates the probabilistic algorithm found in the next subsection.

3.2 Probabilistic Algorithm

In Section 3.2.1, we prepare the lemmas which are used in the algorithms in Sections 3.2.2–3.2.3. We present the probabilistic algorithm with Strategy 1 (Algorithm 2) in Section 3.2.2. We discuss two different strategies (Strategy 2 and Strategy 3) for Algorithm 2 in Section 3.2.3.

3.2.1 Lemmas

We prepare the lemmas which are used in the Sections 3.2.2 and 3.2.3. Lemma 1 is used to linearly transform parameter space. Corollary 1 and Lemma 2 are used to compute the totally degree of \mathcal{D}_{XJ} . Corollary 2 is used in the sampling for interpolation. Lemma 3 is used to compute the degree of \mathcal{D}_{XJ} and to do sampling in Strategy 3. In the statements of Lemmas 1–2 and Corollaries 1–2, we say an affine variety V in \mathbb{C}^n is *non-trivial* if $\mathbb{C}^n \setminus V \neq \emptyset$.

Lemma 1. *For any $G \in \mathbb{Q}[\mathbf{u}]$, there exists a non-trivial affine variety V in \mathbb{C}^n such that for any $(a_1, \dots, a_n) \in \mathbb{C}^n \setminus V$, the total degree of G equals the degree of $B(t_0, t_1, \dots, t_n)$ w.r.t. to t_0 , where*

$$B(t_0, t_1, \dots, t_n) = G(t_0, a_1 t_0 + t_1, \dots, a_n t_0 + t_n)$$

Proof. Suppose the total degree of G is d and G_d is the homogeneous component of G with total degree d . For any $(1, a_1, \dots, a_n) \in \mathbb{C}^{n+1} \setminus \mathcal{V}_a(G_d)$, let $B(t_0, t_1, \dots, t_n) = G(t_0, a_1 t_0 + t_1, \dots, a_n t_0 + t_n)$. It is easily seen that the degree of B w.r.t. t_0 equals d . \square

Corollary 1. *For any $G \in \mathbb{Q}[\mathbf{u}]$, there exists a non-trivial affine variety V in \mathbb{C}^{2n+2} such that for any*

$$(a_0, b_0, \dots, a_n, b_n) \in \mathbb{C}^{2n+2} \setminus V,$$

the total degree of G equals the degree of $B(t)$ where

$$B(t) = G(a_0 t + b_0, \dots, a_n t + b_n).$$

Lemma 2. *There exists a non-trivial affine variety V in \mathbb{C}^{2n+2} such that for any $(a_0, b_0, \dots, a_n, b_n) \in \mathbb{C}^{2n+2} \setminus V$, if*

$$\langle A(t) \rangle = \langle F_0(t), \dots, F_n(t), F_{n+1}, \dots, F_m, J \rangle \cap \mathbb{Q}[t]$$

where $F_i(t)$ is the polynomial by replacing u_i with $a_i t + b_i$ in F_i ($i = 0, \dots, n$) and

$$B(t) = \mathcal{D}_{XJ}(a_0 t + b_0, \dots, a_n t + b_n),$$

then $\langle B(t) \rangle = \sqrt{\langle A(t) \rangle}$.

Proof. By the definition of \mathcal{D}_{XJ} (Notation 5), there exists an affine variety V_1 such that for any $(a_0, b_0, \dots, a_n, b_n) \in \mathbb{C}^{2n+2} \setminus V_1$, $\langle B(t) \rangle$ is radical. Thus, we only need to show that there exists an affine variety V_2 in \mathbb{C}^{2n+2} such that for any $(a_0, b_0, \dots, a_n, b_n) \in \mathbb{C}^{2n+2} \setminus V_2$, $\mathcal{V}_a(\langle B(t) \rangle) = \mathcal{V}_a(\langle A(t) \rangle)$.

Suppose π_t is the canonical projection: $\mathbb{C} \times \mathbb{C}^{m+1} \rightarrow \mathbb{C}$. For any

$$t^* \in \pi_t(\mathcal{V}_a(F_0(t), \dots, F_n(t), F_{n+1}, \dots, F_m, J)),$$

let $u_i^* = a_i t^* + b_i$ (for $i = 0, \dots, n$), then $(u_0^*, \dots, u_n^*) \in \pi(\mathcal{L}_X \cap \mathcal{V}_a(J))$. Hence $\mathcal{D}_{XJ}(u_0^*, \dots, u_n^*) = 0$ and so $B(t^*) = 0$. Thus

$$B(t) \in \mathcal{I}(\pi_t(\mathcal{V}_a(F_0(t), \dots, F_n(t), F_{n+1}, \dots, F_m, J))).$$

Therefore,

$$\begin{aligned} \mathcal{V}_a(A(t)) &= \mathcal{V}_a(\mathcal{I}(\pi_t(\mathcal{V}_a(F_0(t), \dots, F_n(t), F_{n+1}, \dots, F_m, J)))) \\ &\subset \mathcal{V}_a(B(t)). \end{aligned}$$

For any $t^* \in \mathcal{V}_a(\langle B(t) \rangle)$, let $u_i^* = a_i t^* + b_i$ for $i = 0, \dots, n$, then $(u_0^*, \dots, u_n^*) \in \mathcal{V}_a(\mathcal{D}_{XJ}) \subset \mathcal{L}_{XJ}$. By the Extension Theorem [7], there exists an affine variety $V_2 \subset \mathbb{C}^{2n+2}$ such that if $(a_0, b_0, \dots, a_n, b_n) \notin V_2$, then $(u_0^*, \dots, u_n^*) \in \pi(\mathcal{L}_X \cap \mathcal{V}_a(J))$, thus

$$t^* \in \pi_t(\mathcal{V}_a(F_0(t), \dots, F_n(t), F_{n+1}, \dots, F_m, J)) \subset \mathcal{V}_a(A(t)). \square$$

Corollary 2. *There exists a non-trivial affine variety V in \mathbb{C}^n such that for any $(a_1, \dots, a_n) \in \mathbb{C}^n \setminus V$, if*

$$\langle A(u_0) \rangle = \langle F_0, F_1^* \dots, F_n^*, F_{n+1}, \dots, F_m, J \rangle \cap \mathbb{Q}[u_0]$$

where F_i^* is the polynomial by replacing u_i with a_i in F_i ($i = 1, \dots, n$) and

$$B(u_0) = \mathcal{D}_{XJ}(u_0, a_1, \dots, a_n),$$

then $\langle B(u_0) \rangle = \sqrt{\langle A(u_0) \rangle}$.

Lemma 3. *Suppose X is an algebraic model with ML degree N . Let g generate the reduced codimension 1 component of $\langle F_0, \dots, F_m \rangle \cap \mathbb{Q}[u_0, \dots, u_n, y_0]$. If $\text{degree}(g, y_0) = N$ and $D_{y_0} = \text{resultant}(g, \frac{\partial g}{\partial y_0}, y_0)$, then $D_{y_0} \in \langle \mathcal{D}_{XJ} \rangle$.*

Proof. By Notation 5, $\langle \mathcal{D}_{XJ} \rangle$ is radical. So we only need to show $\mathcal{V}(\mathcal{D}_{XJ}) \subset \mathcal{V}(D_{y_0})$. Alternatively, we show $\mathbb{C}^{n+1} \setminus \mathcal{V}(D_{y_0}) \subset \mathbb{C}^{n+1} \setminus \mathcal{V}(\mathcal{D}_{XJ})$. For any $\mathbf{u} \in \mathbb{C}^{n+1} \setminus \mathcal{V}(D_{y_0})$, $g(\mathbf{u}, y_0)$ has N distinct complex solutions w.r.t y_0 . By Projective Extension Theorem [7, Page 403, Chapter 8. Corollary 10], all the N distinct complex solutions can be extended in \mathbb{P}^{m+1} . So the likelihood equations $F_0(\mathbf{u}) = \dots = F_m(\mathbf{u}) = \mathbf{0}$ have N distinct solutions in \mathbb{P}^{m+1} . Therefore, $\mathbf{u} \notin \mathcal{V}(\mathcal{D}_{XJ})$. \square

3.2.2 Probabilistic Algorithm with Strategy 1

We first give an example for the main probabilistic algorithm. We will discuss different ways the main probabilistic algorithm can be adjusted in Section 3.2.3, so we say the main algorithm presented in this subsection is the probabilistic algorithm with Strategy 1.

Example 3 (Algorithm 2 (Strategy 1)). *Suppose the radical of the elimination ideal $\langle F, J \rangle \cap \mathbb{Q}[\mathbf{u}]$ is generated by $D(u_0, u_1, u_2, u_3)$, where $F = u_0 p^3 + u_1 p^2 + u_2 p + u_3$ and $J = \frac{\partial F}{\partial p} = 3u_0 p^2 + 2u_1 p + u_2$. With the standard elimination algorithm, we can compute D to be the unique up to scaling homogeneous polynomial*

$$27u_0^2 u_3^2 - 18u_0 u_1 u_2 u_3 + 4u_0 u_2^3 + 4u_1^3 u_3 - u_1^2 u_2^2.$$

Instead, we will compute D by the steps below.

First, we restrict our parameter space to a general 1-dimensional affine space by replacing each parameter u_i with $\square t + \square$ where \square 's denote a general choice. For example, we substitute $u_0 =$

$7t + 11, u_1 = 3t + 2, u_2 = 5t + 6, u_3 = 4t + 13$ into F and J . We compute the radical of the elimination ideal $\langle F(t, p), J(t, p) \rangle \cap \mathbb{Q}[t]$ to get an ideal generated by

$$t^4 + \frac{173086}{17315}t^3 + \frac{632753}{17315}t^2 + \frac{972374}{17315}t + \frac{531011}{17315}.$$

By Lemma 2 and Corollary 1, the total degree of D equals the degree of the univariate polynomial above, if the 1-dimensional affine space is general (this is because we are intersecting the hypersurface defined by D with a general line). Similarly, we are able to compute the degrees of D with respect to u_0, u_1, u_2, u_3 and get 2, 3, 3, 2, respectively. Since D is only unique up to scaling, to interpolate, we must fix a consistent choice of scaling. If possible, we will fix one of the coefficients of the pure powers $u_i^{\deg D}$ to be 1. However, this is not possible if the coefficients of each pure power is zero, which occurs when there exists no u_i such that the degree of D with respect to u_i equals the total degree. So when this is not possible, we will perform a linear coordinate change so that D has a pure power in the new coordinates with a nonzero coefficient.

Second, we perform a linear change of variables by the following substitution $u_i = \square v_i + \square v_1$ in D to get the polynomial $D^*(v_0, v_1, v_2, v_3)$. For a general substitution, e.g. $u_0 = v_1 - 2v_0, u_1 = v_1, u_2 = v_1 - 3v_2, u_3 = v_1 - 5v_3$, the polynomial D^* has a nonzero coefficient for the pure power $v_1^{\deg D}$.

Now, we may assume D^* has the following structure (here we have a fixed scaling):

$$D^* = v_1^4 + C_1 v_1^3 + C_2 v_1^2 + C_3 v_1 + C_4,$$

where C_i is a homogeneous polynomial in v_0, v_2, v_3 with total degree i . This means the monomials of C_1 are v_0, v_2, v_3 , and we have $C_1 = C_{11}v_0 + C_{12}v_2 + C_{13}v_3$. Once we determine the coefficients C_{11}, C_{12}, C_{13} , we are done with C_1 . This will be done by interpolation; rather than determining D^* directly, we determine D^* restricted to a line which induces linear constraints on the coefficients of D^* . For example, we substitute $v_0 = 13, v_2 = 4, v_3 = 5$ into $F^*(v_0, v_1, v_2, v_3, p)$ and $J^*(v_0, v_1, v_2, v_3, p)$, and then we compute the radical of the elimination ideal $\langle F^*(13, v_1, 4, 5, p), J^*(13, v_1, 4, 5, p) \rangle \cap \mathbb{Q}[v_1]$ finding it is generated by

$$v_1^4 + \frac{243}{2}v_1^3 + \frac{87939}{16}v_1^2 + \frac{425385}{4}v_1 + \frac{2896803}{4}. \quad (1)$$

By Corollary 2, the polynomial in (1) is $D^*(13, v_1, 4, 5)$. Similarly, we compute to find

$$D^*(7, v_1, 3, 11) = v_1^4 - 168v_1^3 + \frac{36873}{4}v_1^2 - \frac{690201}{4}v_1 + \frac{4012281}{4} \quad (2)$$

$$D^*(2, v_1, 8, 9) = v_1^4 - \frac{221}{2}v_1^3 + \frac{57627}{16}v_1^2 - \frac{60183}{2}v_1 + 68499. \quad (3)$$

Comparing the coefficients of D^* to Equations (1–3), we have the linear relations below

$$13C_{11} + 4C_{12} + 5C_{13} = -\frac{243}{2}, \quad 7C_{11} + 3C_{12} + 11C_{13} = -168, \quad 2C_{11} + 8C_{12} + 9C_{13} = -\frac{221}{2}.$$

Solving the linear system yields $C_{11} = -5, C_{12} = \frac{3}{2}, C_{13} = -\frac{25}{2}$. We similarly compute C_2, C_3, C_4 to determine D^* :

$$\begin{aligned} & \frac{27}{4}v_0^2v_1^2 - \frac{135}{2}v_0^2v_1v_3 + \frac{675}{4}v_0^2v_3^2 - 5v_0v_1^3 - \frac{9}{2}v_0v_1^2v_2 + \frac{225}{4}v_0v_1^2v_3 + \\ & -\frac{27}{2}v_0v_1v_2^2 + \frac{135}{4}v_0v_1v_2v_3 - \frac{675}{4}v_0v_1v_3^2 + \frac{27}{2}v_0v_2^3 + v_1^4 + \frac{3}{2}v_1^3v_2 + \\ & -\frac{25}{2}v_1^3v_3 + \frac{99}{16}v_1^2v_2^2 - \frac{135}{8}v_1^2v_2v_3 + \frac{675}{16}v_1^2v_3^2 - \frac{27}{4}v_1v_2^3. \end{aligned} \quad (4)$$

By applying the inverse linear change of coordinates $v_0 = -\frac{u_0}{2} + \frac{u_1}{2}, v_1 = u_1, v_2 = -\frac{u_2}{3} + \frac{u_1}{3}, v_3 = -\frac{u_3}{5} + \frac{u_1}{5}$ to D^* and removing the denominator, we recover D .

We present the formal description of the probabilistic algorithm in Algorithm 2. We explain the main algorithm (Algorithm 2) and all the sub-algorithms (Algorithms 3–5) below.

Algorithm 5 (Degree). The probabilistic algorithm terminates correctly by Corollary 1 and Lemma 2.

Algorithm 4 (LinearOperator). The probabilistic algorithm terminates correctly by Lemma 1.

Algorithm 3 (Intersect). The probabilistic algorithm terminates correctly by Corollary 2.

Algorithm 2 (InterpolationDX-J).

Lines 1–5. We compute the total degree of \mathcal{D}_{XJ} and the degrees of \mathcal{D}_{XJ} with respect to u_0, \dots, u_d : d, d_0, \dots, d_n by Algorithm 5. Algorithm 4 guarantees that $d_0 = d$ by applying a proper linear transformation $u_1 = a_1 \cdot u_0 + u_1, \dots, u_n = a_n \cdot u_0 + u_n$.

Lines 6–7. Suppose $\mathcal{D}_{XJ} = u_0^d + C_1 u_0^{d-1} + \dots + C_{d-1} u_0 + C_d$ where $C_1, \dots, C_d \in \mathbb{Q}[u_1, \dots, u_n]$ and the total degree of C_j is j . For $j = 1, \dots, n$, we estimate all the possible monomials of C_j by computing the set

$$\{u_1^{\alpha_1} \dots u_n^{\alpha_n} \mid \alpha_1 + \dots + \alpha_n = j, 0 \leq \alpha_i \leq d_i\}$$

Assume the cardinality of the set is N_j and rename these monomials as $U_{j,1}, \dots, U_{j,N_j}$. Then we assume

$$C_j = c_{j,1} U_{j,1} + \dots + c_{j,N_j} U_{j,N_j}$$

where $c_{j,1}, \dots, c_{j,N_j} \in \mathbb{Q}$. The rest of the algorithm is to compute $c_{j,1}, \dots, c_{j,N_j}$.

Lines 8–12. For each j , for $k = 1, \dots, N_j$, for a random integer vector $\mathbf{b}_k = (b_{k,1}, \dots, b_{k,n})$, we compute $\mathcal{D}_{XJ}(u_0, \mathbf{b}_k)$ by Algorithm 3. That means to compute the function value $C_j(\mathbf{b}_k)$ without knowing \mathcal{D}_{XJ} .

Lines 13–15. For each j , we solve a square linear equation system for the unknowns $c_{j,1}, \dots, c_{j,N_j}$:

$$\begin{aligned} c_{j,1} U_{j,1}(\mathbf{b}_k) + \dots + c_{j,N_j} U_{j,N_j}(\mathbf{b}_k) &= C_j(\mathbf{b}_k), \\ (k &= 1, \dots, N_j) \end{aligned}$$

It is known that we can choose nice \mathbf{b}_k probabilistically such that the coefficient matrix of the linear equation system is non-singular.

Lines 16. We apply the inverse linear transformation in the parameter space to get the \mathcal{D}_{XJ} for the original F_0, \dots, F_m .

We can also apply the interpolation idea to Algorithm PROPERNESSDEFECTS [25] and get a probabilistic algorithm to compute the $\mathcal{D}_{X\infty}$. We omit the formal description of the algorithm.

Remark 3. According to the Notation 5, when the codimension of $\mathcal{L}_{XJ} \setminus \{\mathbf{0}\}$ ($\mathcal{L}_{X\infty} \setminus \{\mathbf{0}\}$) is greater than 1, we define \mathcal{D}_{XJ} ($\mathcal{D}_{X\infty}$) is 1. In this case, the number of real/positive solutions remains constant over a dense Zariski open set of the entire parameter space.

3.2.3 Strategy 2 and Strategy 3

In this section, we discuss different strategies for the probabilistic algorithm of Section 3.2.2. Note in Algorithm 2, lines 6–9, we compute all the possible terms in \mathcal{D}_{XJ} and in Algorithm 2, lines 8–15, we interpolate all these terms at once. From the timings in the “Algorithm 2-Strategy 1” columns of Tables 1-2, we see for most of examples, Strategy 1 does not perform better than Algorithm 1.

By experiments, we find the main problem of Strategy 1 is the lifting step. In fact, it is expensive to compute the inverse of \mathcal{M}_j in Algorithm 2, Line 14, which can be a large size matrix with (large) rational entries. In order to overcome this problem, we interpolate one parameter by one parameter, say Strategy 2. We omit the naive formal description of the algorithm, instead to explain how Strategy 2 works, we provide Example 4.

Example 4 (Algorithm 2 (Strategy 2)). We compute D in Example 3 by Strategy 2. Since the first two steps of Strategy 2 are the same as that of Strategy 1 (outlined in Example 3), we only show how to compute D^* . Recall D^* is a homogeneous polynomial in v_0, v_1, v_2, v_3 with total degree 4 and the degrees of D^* with respect to v_0, v_1, v_2, v_3 are 2, 4, 3, 2.

Algorithm 2: (Main Algorithm) InterpolationDX-J

```

input :  $F_0, \dots, F_m, J$ 
output:  $\mathcal{D}_{XJ}$ 
1  $a_1, \dots, a_n \leftarrow \text{LinearOperator}(F_0, \dots, F_m, J)$ 
2 for  $i$  from 1 to  $n$  do
3    $F'_i \leftarrow$  replace  $u_i$  in  $F_i$  with  $a_i u_0 + u_i$ 
4  $NewSys \leftarrow F_0, F'_1, \dots, F'_n, F_{n+1}, \dots, F_m, J$ 
5  $d, d_0, \dots, d_n \leftarrow \text{Degree}(NewSys)$ 
6 for  $j$  from 1 to  $d$  do
7   Rename all the monomials of the set
      
$$\{u_1^{\alpha_1} \dots u_n^{\alpha_n} \mid \alpha_1 + \dots + \alpha_n = j, 0 \leq \alpha_i \leq d_i\}$$

      as  $U_{j,1}, \dots, U_{j,N_j}$ 
8  $N \leftarrow \max(N_1, \dots, N_d)$ 
9 for  $k$  from 1 to  $N$  do
10   $b_{k,1}, \dots, b_{k,n} \leftarrow$  random integers
11   $A(u_0) \leftarrow \text{Intersect}(NewSys, b_{k,1}, \dots, b_{k,n})$ 
12   $C_{d,k}^*, \dots, C_{1,k}^* \leftarrow$  the coefficients of  $A(u_0)$  with respect to  $u_0^0, \dots, u_0^{d-1}$ 
13 for  $j$  from 1 to  $d$  do
14   $\mathcal{M}_j \leftarrow N_j \times N_j$  matrix whose  $(k, r)$ -entry is  $U_{j,r}(b_{k,1}, \dots, b_{k,n})$ 
15   $C_j \leftarrow (U_{j,1}, \dots, U_{j,N_j}) \mathcal{M}_j^{-1} (C_{j,1}^*, \dots, C_{j,N_j}^*)^T$ 
16  $\mathcal{D}_{XJ} \leftarrow$  replace  $u_1, \dots, u_n$  in  $u_0^d + \sum_{i=0}^{d-1} C_{d-i} u_0^i$  with  $u_1 - a_1 \cdot u_0, \dots, u_n - a_n \cdot u_0$ 
17 Return  $\mathcal{D}_{XJ}$ 

```

Algorithm 3: Intersect

```

input :  $F_0, \dots, F_m, J$  and integers  $b_1, \dots, b_n$ 
output:  $\mathcal{D}_{XJ}(u_0, b_1, \dots, b_n)$ 
1 for  $i$  from 1 to  $n$  do
2    $F_i^* \leftarrow$  replace  $u_i$  in  $F_i$  with  $b_i$ 
3  $A(u_0) \leftarrow$  generator of the radical of elimination ideal  $\langle F_0, F_1^*, \dots, F_n^*, F_{n+1}, \dots, F_m, J \rangle \cap \mathbb{Q}[u_0]$ 
4 return  $A(u_0)$ 

```

Algorithm 4: LinearOperator

```

input :  $F_0, \dots, F_m, J$ 
output:  $a_1, \dots, a_n$  such that the total degree of  $\mathcal{D}_{XJ}$  equals the degree of  $\mathcal{D}_{XJ}(u_0, a_1 \cdot u_0 + u_1, \dots, a_n \cdot u_0 + u_n)$ 
with respect to  $u_0$ 
1  $d, d_0, \dots, d_n \leftarrow \text{Degree}(F_0, \dots, F_m, J)$ 
2 if  $d = d_0$  then
3   return  $0, \dots, 0$ 
4 else
5   for  $i$  from 1 to  $n$  do
6      $a_i \leftarrow$  a random integer
7      $F'_i \leftarrow$  replace  $u_i$  in  $F_i$  with  $a_i \cdot u_0 + u_i$ 
8    $NewSys \leftarrow F_0, F'_1, \dots, F'_n, F_{n+1}, \dots, F_m, J$ 
9    $d, d_0, \dots, d_n \leftarrow \text{Degree}(NewSys)$ 
10 return  $a_1, \dots, a_n$ 

```

Algorithm 5: Degree

```

input :  $F_0, \dots, F_m, J$ 
output:  $d, d_0, \dots, d_n$ , where  $d$  is the total degree of  $\mathcal{D}_{XJ}$  and  $d_i$  is the degree of  $\mathcal{D}_{XJ}$  with respect to each  $u_i$ 
1 for  $i$  from 0 to  $n$  do
2    $F_0^*, \dots, F_n^* \leftarrow$  replace  $u_0, \dots, u_{i-1}, u_{i+1}, \dots, u_n$  in  $F_0, \dots, F_n$  with random integers
3    $A(u_i) \leftarrow$  generator of the radical of elimination ideal  $\langle F_0^*, \dots, F_n^*, F_{n+1}, \dots, F_m, J \rangle \cap \mathbb{Q}[u_i]$ 
4    $d_i \leftarrow$  degree of  $A(u_i)$ 
5    $a_i, b_i \leftarrow$  random integers
6  $F_0(t), \dots, F_n(t) \leftarrow$  replace  $u_0, \dots, u_n$  with  $a_0 \cdot t + b_0, \dots, a_n \cdot t + b_n$  in  $F_0, \dots, F_n$ 
7  $A(t) \leftarrow$  generator of the radical of elimination ideal  $\langle F_0(t), \dots, F_n(t), F_{n+1}, \dots, F_m, J \rangle \cap \mathbb{Q}[t]$ 
8  $d \leftarrow$  degree of  $A(t)$ 
9 return  $d, d_0, \dots, d_n$ 

```

First, replace all but two parameters with general numbers, e.g. $v_2 = 3$ and $v_3 = 5$, and we will compute $D^*(v_0, v_1, 3, 5)$. We may assume $D^*(v_0, v_1, 3, 5)$ has the following form

$$D^*(v_0, v_1, 3, 5) = v_1^4 + C_1 v_1^3 + C_2 v_1^2 + C_3 v_1 + C_4,$$

where C_i ($i = 1, 2, 3, 4$) is a univariate polynomial in v_0 . The degree of C_i is bounded above by i and the degree of D^* with respect to v_0 . For instance, the degree of C_4 is at most 2, so we may assume

$$C_4 = C_{40}v_0^2 + C_{41}v_0 + C_{42},$$

where the coefficients C_{40}, C_{41}, C_{42} are constants because v_2 and v_3 have been fixed. We determine these coefficients by interpolation. Now, we continue to have $v_2 = 3, v_3 = 5$, and also fix v_0 to be a general number, e.g. $v_0 = 2$, to determine the elimination ideal

$$\langle F(2, v_1, 3, 5, p), J(2, v_1, 3, 5, p) \rangle \cap \mathbb{Q}[v_1] = \langle v_1^4 - 68v_1^3 + \frac{5733}{4}v_1^2 - \frac{36801}{4}v_1 + 17604 \rangle.$$

By Corollary 2,

$$D^*(2, v_1, 3, 5) = v_1^4 - 68v_1^3 + \frac{5733}{4}v_1^2 - \frac{36801}{4}v_1 + 17604.$$

Similarly, by fixing v_0 to be different values, we also determine

$$\begin{aligned} D^*(7, v_1, 3, 5) &= v_1^4 - 93v_1^3 + \frac{6219}{2}v_1^2 - \frac{174231}{4}v_1 + \frac{837081}{4} \\ D^*(9, v_1, 3, 5) &= v_1^4 - 103v_1^3 + \frac{7749}{2}v_1^2 - \frac{248103}{4}v_1 + \frac{1379997}{4}. \end{aligned}$$

These 3 univariate polynomials induce linear constraints on the coefficients C_{40}, C_{41}, C_{42} :

$$2^2 C_{40} + 2C_{41} + C_{42} = 17604, \quad 7^2 C_{40} + 7C_{41} + C_{42} = \frac{837081}{4}, \quad 9^2 C_{40} + 9C_{41} + C_{42} = \frac{1379997}{4}.$$

Solving this linear system, we find $C_4 = \frac{16875}{4}v_0^2 + \frac{729}{2}v_0$.

Similarly, we compute C_1, C_2, C_3 and get $D^*(v_0, v_1, 3, 5)$: $v_1^4 + \frac{27}{4}v_0^2v_1^2 - \frac{675}{2}v_0^2v_1 + \frac{16875}{4}v_0^2 - 5v_0v_1^3 + \frac{549}{2}v_0v_1^2 - 3834v_0v_1 + \frac{729}{2}v_0 - 58v_1^3 + \frac{3429}{4}v_1^2 - \frac{729}{4}v_1$.

Second, we compute $D^*(v_0, v_1, v_2, 5)$. According to the monomials of $D^*(v_0, v_1, 3, 5)$, we may assume $D^*(v_0, v_1, v_2, 5)$ has the following form:

$$v_1^4 + C_1 v_0^2 v_1^2 + C_2 v_0^2 v_1 + C_3 v_0^2 + C_4 v_0 v_1^3 + \cdots + C_6 v_0 v_1 + C_7 v_0 + C_8 v_1^3 + C_9 v_1^2 + C_{10} v_1,$$

where C_i ($i = 1, \dots, 10$) is a polynomial in v_2 . The degree of C_i is bounded above by the degree of D^* with respect to v_2 and the difference $\deg(D^*) - \deg(T_i)$ where T_i corresponds to the term of C_i in $D^*(v_0, v_1, v_2, 5)$. For instance, the degree of C_{10} is 3, so

$$C_{10} = C_{100}v_2^3 + C_{101}v_2^2 + C_{102}v_2 + C_{103}.$$

Now we solve the coefficients $C_{100}, C_{101}, C_{102}, C_{103}$. For general specializations $v_2 = 11, 21, 4$, we repeat the first step and compute

$$D^*(v_0, v_1, 11, 5) = \frac{27}{4}v_0^2v_1^2 - \frac{675}{2}v_0^2v_1 + \frac{16875}{4}v_0^2 - 5v_0v_1^3 + \frac{513}{2}v_0v_1^2 - 3996v_0v_1 + \frac{35937}{2}v_0 + v_1^4 - 46v_1^3 + \frac{3501}{4}v_1^2 - \frac{35937}{4}v_1.$$

$$D^*(v_0, v_1, 21, 5) = \frac{27}{4}v_0^2v_1^2 - \frac{675}{2}v_0^2v_1 + \frac{16875}{4}v_0^2 - 5v_0v_1^3 + 234v_0v_1^2 - \frac{13757}{2}v_0v_1 - \frac{250047}{2}v_0 + v_1^4 - 31v_1^3 + \frac{4023}{2}v_1^2 - \frac{250047}{4}v_1.$$

$$D^*(v_0, v_1, 4, 5) = \frac{27}{4}v_0^2v_1^2 - \frac{675}{2}v_0^2v_1 + \frac{16875}{4}v_0^2 - 5v_0v_1^3 - \frac{1089}{4}v_0v_1^2 - \frac{15039}{4}v_0v_1 - 864v_0 + v_1^4 - \frac{113}{2}v_1^3 + \frac{13059}{16}v_1^2 - 432v_1.$$

These bivariate polynomials induce linear constraints on $C_{100}, C_{101}, C_{102}, C_{103}$:

$$3^3 C_{100} + 3^2 C_{101} + 3C_{102} + C_{103} = -\frac{729}{4}$$

$$11^3 C_{100} + 11^2 C_{101} + 11C_{102} + C_{103} = -\frac{35937}{4}$$

$$21^3 C_{100} + 21^2 C_{101} + 21 C_{102} + C_{103} = -\frac{250047}{4}$$

$$4^3 C_{100} + 4^2 C_{101} + 4 C_{102} + C_{103} = -432.$$

Solving this system determines C_{10} . Similarly, we find C_1, \dots, C_9 to get $D^*(v_0, v_1, v_2, 5)$:

$$\frac{27}{4}v_0^2v_1^2 - \frac{675}{2}v_0^2v_1 + \frac{16875}{4}v_0^2 - 5v_0v_1^3 - \frac{9}{4}v_0v_1^2v_2 + \frac{1125}{4}v_0v_1^2 - \frac{27}{2}v_0v_1v_2^2 + \frac{675}{4}v_0v_1v_2 - \frac{16875}{4}v_0v_1 + \frac{27}{2}v_0v_2^3 + v_1^4 + \frac{3}{2}v_1^3v_2 - \frac{125}{2}v_1^3 + \frac{99}{16}v_1^2v_2^2 - \frac{675}{8}v_1^2v_2 + \frac{16875}{16}v_1^2 - \frac{27}{4}v_1v_2^3.$$

Third, it is easy to recover D^* from $D^*(v_0, v_1, v_2, 5)$ because we know D^* is homogeneous. For instance, the second term in $D^*(v_0, v_1, v_2, 5)$ is $-\frac{675}{2}v_0^2v_1$. The total degree of this term is 3. So the corresponding term in D^* is $\frac{-\frac{675}{2}v_0^2v_1v_3}{5} = -\frac{135}{2}v_0^2v_1v_3$. Similarly, we recover all terms of D^* and again get (4).

From the timings in the columns “Algorithm 2. Strategy 2” of Tables 1–2, we see Strategy 2 improves the efficiency for the larger examples, e.g. the Dense Models 3–4 and Models 3–4. The advantage of Strategy 2 compared to Strategy 1 is that we only need to solve a small-sized (bounded by the degree of $\mathcal{D}_{X,J}$ with respect to each parameter at each step) linear equation system in the lifting step. Therefore, the most time consuming part of Strategy 2 is sampling. As soon as we can compute the degree of $\mathcal{D}_{X,J}$, we can estimate the computational timings of Strategy 2. In fact, we can first check the approximate timing of sampling once by running Algorithm 3 once, say T_s . Suppose the total degree of $\mathcal{D}_{X,J}$ is d and the degrees with respect to u_0, \dots, u_n are d_0, \dots, d_n . Without loss of generality, assume $d = d_0$ and $d_0 \geq d_1 \geq \dots \geq d_n$. Then the total approximate sampling timing of Strategy 2 is

$$T_s d_2 \cdots d_n. \quad (5)$$

So if we improve T_s , we might improve the efficiency. This motivates Strategy 3.

We give another two sub-algorithms Algorithm 6 and Algorithm 7 to do sampling for interpolation and to compute the degree of $\mathcal{D}_{X,J}$, respectively. More precisely, Strategy 3 means to use Algorithms 6 and 7 instead of Algorithms 3 and 5 in Algorithm 2. The correctness of Algorithms 6 and 7 directly follows from Lemma 3. We give Example 5 to show how Algorithm 7 works (and similarly this example shows how Algorithm 6 works).

Example 5 (Toy Example for Strategy 3). *Here, we illustrate Algorithm 7 which is used to compute the degree of the discriminant. Here, we compute the total degree of $\mathcal{D}_{X,J}$ to be 4 of the linear model in Example 2 by Algorithm 7. Define F_0, F_1, \dots, F_5 as those in Example 1. Rename $p_0, \dots, p_3, \lambda_1, \lambda_2$ as $y_0, \dots, y_3, y_4, y_5$. The determinant of Jacobian matrix of F_0, F_1, \dots, F_5 with respect to $y_0, \dots, y_3, y_4, y_5$, say J , is*

$$\begin{aligned} & -\lambda_1^2 p_0 p_1 - 4\lambda_1^2 p_0 p_2 - 25\lambda_1^2 p_0 p_3 - \lambda_1^2 p_1 p_2 - 36\lambda_1^2 p_1 p_3 - 49\lambda_1^2 p_2 p_3 + \lambda_1 \lambda_2 p_0 p_1 \\ & + 8\lambda_1 \lambda_2 p_0 p_2 - 125\lambda_1 \lambda_2 p_0 p_3 + 3\lambda_1 \lambda_2 p_1 p_2 - 144\lambda_1 \lambda_2 p_1 p_3 - 147\lambda_1 \lambda_2 p_2 p_3 \\ & + 12\lambda_2^2 p_0 p_1 + 32\lambda_2^2 p_0 p_2 - 150\lambda_2^2 p_0 p_3 + 4\lambda_2^2 p_1 p_2 - 108\lambda_2^2 p_1 p_3 - 98\lambda_2^2 p_2 p_3. \end{aligned}$$

First, we restrict the parameter space to a general line by the substitution $u_i = \square t + \square$, e.g. $u_0 = t + 13, u_1 = 4t + 2, u_2 = 9t + 6, u_3 = 31t + 5$, into F_0, F_1, \dots, F_5 . Next, we eliminate all but one unknown; we compute the radical of the elimination ideal $\langle F_0(t, y_0, \dots, y_6), \dots, F_5(t, y_0, \dots, y_6) \rangle \cap \mathbb{Q}[t, y_0]$, say $\langle g \rangle$ where g is

$$\begin{aligned} & 20250y_0^3t^2 + 23400y_0^3t - 22770y_0^2t^2 + 6760y_0^3 - 45961y_0^2t + \\ & 1488y_0t^2 - 18954y_0^2 + 20582y_0t - 24t^2 + 16094y_0 - 624t - 4056. \end{aligned}$$

Computing the resultant $\text{resultant}(g, \frac{\partial g}{\partial y_0}, y_0)$, we get $(t + 13)(t + \frac{45}{26})G$, where

$$G = t^4 + \frac{34880663}{24297438}t^3 + \frac{1144507225}{1166277024}t^2 + \frac{391327027}{1166277024}t + \frac{244617385}{4665108096}.$$

By Lemma 3, we know $\mathcal{D}_{X,J}(t + 13, 4t + 2, 9t + 6, 31t + 5)$ is a factor of $\text{resultant}(g, \frac{\partial g}{\partial y_0}, y_0)$. For the three irreducible factors of $\text{resultant}(g, \frac{\partial g}{\partial y_0}, y_0)$, we check by computing elimination ideal that

$$\langle F_0(t, y_0, \dots, y_6), \dots, F_5(t, y_0, \dots, y_6), J, t + 13 \rangle \cap \mathbb{Q}[t] = \langle 1 \rangle$$

$$\langle F_0(t, y_0, \dots, y_6), \dots, F_5(t, y_0, \dots, y_6), J, t + \frac{45}{26} \rangle \cap \mathbb{Q}[t] = \langle 1 \rangle$$

$$\langle F_0(t, y_0, \dots, y_6), \dots, F_5(t, y_0, \dots, y_6), J, G \rangle \cap \mathbb{Q}[t] = \langle G \rangle.$$

Therefore we confirm $\mathcal{D}_{X,J}(t + 13, 4t + 2, 9t + 6, 31t + 5)$ is G , which is consistent with the result computed by Algorithm 5. Again, we conclude the total degree of $\mathcal{D}_{X,J}$ is 4.

In Algorithm 5 we are eliminating all unknowns but t in $\langle F_0, \dots, F_m, J \rangle$. While in Algorithm 7 we are eliminating all unknowns but t, y_0 in $\langle F_0, \dots, F_m \rangle$ and carrying out many other eliminations to t in ideals $\langle F_0, \dots, F_m, J, G_i \rangle$. It is not immediately clear why Algorithm 7 would be an improvement over Algorithm 5. In fact, for the larger algebraic models, J will be a huge polynomial. In this case, eliminating from $\langle F_0, \dots, F_m, J \rangle$ can be very slow. Either eliminating from $\langle F_0, \dots, F_m \rangle$ or eliminating from $\langle F_0, \dots, F_m, J, G_i \rangle$ can be faster because either removing huge polynomials or including additional small polynomials might speed up elimination according to experience. Of course it is difficult to prove theoretically which strategy will be more efficient, the best way is to implement both strategies and run the benchmarks. In our case, Algorithm 7 indeed gives improvement for some larger models, see Models 4, 6, 7, 9 in Table 3, Section 4.

Algorithm 6: IntersectStrategy3

```

input :  $F_0, \dots, F_m, J$  and integers  $b_1, \dots, b_n$ 
output:  $\mathcal{D}_{X,J}(u_0, b_1, \dots, b_n)$ 
1 for  $i$  from 1 to  $n$  do
2    $F_i^* \leftarrow$  replace  $u_i$  in  $F_i$  with  $b_i$ 
3  $A_1(u_0, y_0) \leftarrow$  generator of the radical of elimination ideal  $\langle F_0^*, \dots, F_n^*, F_{n+1}, \dots, F_m \rangle \cap \mathbb{Q}[u_0, y_0]$ 
4 if  $\text{degree}(A_1, y_0) = \text{ML Degree}$  then
5    $A(u_0) \leftarrow \text{resultant}(A_1, \frac{\partial A_1}{\partial y_0}, y_0)$ 
6  $\text{newA} \leftarrow 1$ 
7 for each irreducible factor  $G(u_0)$  of  $A(u_0)$  do
8   if  $\langle F_0^*, \dots, F_n^*, F_{n+1}, \dots, F_m, J, G(u_0) \rangle \cap \mathbb{Q}[u_0] \neq \langle 1 \rangle$  then
9      $\text{newA} \leftarrow \text{newA} \cdot G(u_0)$ 
10 return  $\text{newA}$ 

```

Remark 4. Strategy 3 is incomplete whenever the solutions to the likelihood equations along the general line do not have distinct y_0 coordinates. This means the “if” condition in Algorithm 6–Line 4 or in Algorithm 7–Lines 4, 14 is not satisfied. In principle, we can modify Strategy 3 by applying a general linear coordinate change to y_0, \dots, y_m .

4 Implementation Details and Experimental Timings

We have implemented the probabilistic method, Algorithm 2 in Maple 2015. For comparisons, we have also implemented the standard Algorithm 1 in Maple 2015. We run our implementations of Algorithms 1 and 2 for many examples to set benchmarks by a 3.2 GHz Inter Core i5 processor (8GB total memory) under OS X 10.9.3¹. We record the timings for these examples in Tables 1–3². In this section, we give conclusions from our computational experiments and results.

The probabilistic algorithm is implemented in three different strategies described in Section 3.2.3; they are the ‘interpolate every term at once’ strategy; ‘interpolate one parameter at a time’ strategy; and ‘one variable’ strategy respectively.

There are two kinds of benchmarks, the random models and literature models. The random models, say Dense Models 1–6, are dense homogeneous polynomials generated by the Maple command

$$\text{randpoly}([p_0, \dots, p_n], \text{dense}, \text{degree} = m), \quad n = 2, 3; \quad m = 2, 3, 4.$$

¹See our Maple code on the website: <https://sites.google.com/site/rootclassification/publications/jsc2016>

²When a computation returned no output in 2 hours we record “>2h”

Algorithm 7: DegreeStrategy3

```

input :  $F_0, \dots, F_m, J$ 
output:  $d, d_0, \dots, d_n$ , where  $d$  is the total degree of  $\mathcal{D}_{X,J}$  and  $d_i$  is the degree of  $\mathcal{D}_{X,J}$  w.r.t each  $u_i$  ( $i = 0, \dots, n$ )
1 for  $i$  from 0 to  $n$  do
2    $F_0^*, \dots, F_n^* \leftarrow$  replace  $u_0, \dots, u_{i-1}, u_{i+1}, \dots, u_n$  in  $F_0, \dots, F_n$  with random integers
3    $g(u_i, y_0) \leftarrow$  generator of the radical of elimination ideal  $\langle F_0^*, \dots, F_n^*, F_{n+1}, \dots, F_m \rangle \cap \mathbb{Q}[u_i, y_0]$ 
4   if  $\text{degree}(g, y_0) = \text{ML Degree}$  then
5      $A(u_i) \leftarrow \text{resultant}(g, \frac{\partial g}{\partial y_0}, y_0)$ 
6    $\text{newA} \leftarrow 1$ 
7   for each irreducible factor  $G(u_i)$  of  $A(u_i)$  do
8     if  $\langle F_0^*, \dots, F_n^*, F_{n+1}, \dots, F_m, J, G(u_i) \rangle \cap \mathbb{Q}[u_i] \neq \langle 1 \rangle$  then
9        $\text{newA} \leftarrow \text{newA} \cdot G(u_i)$ 
10   $d_i \leftarrow$  degree of  $\text{newA}$ 
11   $a_i, b_i \leftarrow$  random integers
12   $F_0(t), \dots, F_n(t) \leftarrow$  replace  $u_0, \dots, u_n$  with  $a_0 \cdot t + b_0, \dots, a_n \cdot t + b_n$  in  $F_0, \dots, F_n$ 
13   $g(t, y_0) \leftarrow$  generator of the radical of elimination ideal  $\langle F_0(t), \dots, F_n(t), F_{n+1}, \dots, F_m \rangle \cap \mathbb{Q}[t, y_0]$ 
14  if  $\text{degree}(g, y_0) = \text{ML Degree}$  then
15     $A(t) \leftarrow \text{resultant}(g, \frac{\partial g}{\partial y_0}, y_0)$ 
16   $\text{newA} \leftarrow 1$ 
17  for each irreducible factor  $G(t)$  of  $A(t)$  do
18    if  $\langle F_0^*, \dots, F_n^*, F_{n+1}, \dots, F_m, J, G(t) \rangle \cap \mathbb{Q}[t] \neq \langle 1 \rangle$  then
19       $\text{newA} \leftarrow \text{newA} \cdot G(t)$ 
20   $d \leftarrow$  degree of  $\text{newA}$ 
21 return  $d, d_0, \dots, d_n$ 

```

Models 1–9 are examples presented in the literatures [19, 9, 17]. See the appendix for the literature models' defining equations³.

The most important conclusion is that our method with new implementation based on **Maple** 2015 and **FGb** is more efficient than our implementation in [27]. For the random models, in [27, Table 1], our previous implementation of Algorithm 2 was only able to compute random dense models with $n + 1 = 3$ and degree of $X \leq 3$. With the new implementation, we have dramatic speed ups as shown by the second and third row of Table 1. Furthermore, we can compute larger dense models, see the fourth and fifth row of the Table 1.

In addition, we see in Table 2 that the new implementation has improved performance for literature models as well. Models 1–4 are exactly the same four literature models in [27, Table 2]. For instance, for Model 4 [27, Example 6], the old implementation of Strategy 2 spent 30 days while the new implementation of Strategy 2 takes less than 30 minutes.

The probabilistic algorithm performs better for the larger sized models (Dense Models 3–4, Models 3–4) than the standard method (the standard method performed better for the smaller models). For Models 4–9, while using the standard algorithm our computer reaches its memory limits after running for 2–3 hours. One explanation for this is seen in Table 4. We see the intermediate matrices generated by Algorithm 1 using **FGb** are quite large.

For most larger literature models, the new Strategy 3 improves the sampling timings compared to Strategy 2 (see Models 4, 6, 7, 9 in Table 3).

It is a challenging task to determine the discriminants for Models 5–9 by the probabilistic algorithm proposed in this paper (interpolation based on linear lifting). One way to improve the methods is to take advantage of sparsity of the monomial support of the discriminant. For example, the discriminant for Model 4 has only 1307 terms out of the possible 6188.

Another important take away from the tables is that the probabilistic algorithm with Strategy 1 is comparable to the standard elimination method in Table 2. We only receive great improvements when employing Strategies 2 or 3.

Example 6. *As a guide to reading the tables, we summarize the first row of Table 1. The model*

³The dense models are available at <https://sites.google.com/site/rootclassification/publications/jsc2016>

invariant of Dense Model 1 in Table 1 is

$$-37p_0^2 - 68p_0p_1 - 64p_0p_2 + 26p_1^2 + 18p_1p_2 + 20p_2^2.$$

We see the number of probability variables (p_0, p_1, p_2) is 3 and the total degree of this model invariant is 2. It is straightforward to check the ML degree is 6. By Algorithm 5 or 7, we compute the total degree of $\mathcal{D}_{X,J}$ is 10. Note that when we could not compute the total degree of $\mathcal{D}_{X,J}$ we record “?”.

Table 1: Computing $\mathcal{D}_{X,J}$ for random dense models (s: seconds; h: hours).

Dense Models	# p_i	degree X	ML degree X	degree $\mathcal{D}_{X,J}$	Algorithm 1	Algorithm 2 (Interpolation)	
					(Elimination)	Strategy 1	Strategy 2
Dense Model 1	3	2	6	10	(4.18s) 0.04s	(.76s) 0.4s	(.61s) 0.4s
Dense Model 2	3	3	12	24	1.2s	(1107s) 2.9s	(1180s) 2.5s
Dense Model 3	3	4	20	44	46.4s	27.3s	21.8s
Dense Model 4	4	2	14	34	3394.6s	>2h	252.0s
Dense Model 5	4	3	39	114	>2h	>2h	>2h
Dense Model 6	4	4	84	?	>2h	>2h	>2h

Table 1 provides the timings of Algorithm 1 and Algorithm 2 (with Strategies 1–2) for Dense Models 1–6. The columns “# p_i ”, “degree of X ”, “ML degree X ” and “degree $\mathcal{D}_{X,J}$ ” give the model information: number of probability variables, total degree of the model invariant, ML degree, and total degree of $\mathcal{D}_{X,J}$ respectively. Timings in parenthesis are the average timings from our previous implementation.

Table 2: Computing $\mathcal{D}_{X,J}$ for literature models (s: seconds; h: hours; d: days)

Models	ML degree X	degree $\mathcal{D}_{X,J}$	Algorithm 1	Algorithm 2 (Interpolation)	
			(Elimination)	Strategy 1	Strategy 2
Model 1	3	6	(11.1s) 0.23s	(5.3s) 1.4s	(6.4s) 1.7s
Model 2	2	4	(36446s) 4.4s	(360s) 2.4s	(56.3s) 3.3s
Model 3	4	14	(>16h) 4242.3s	(>16h) >2h	(2768s) 370.3s
Model 4	6	12	(>12d) >2h	(>30d) >2h	(30d) 1501.0s

Table 2 provides the timings of Algorithm 1 and Algorithm 2 (with Strategies 1–2) for Models 1–4. Timings in parenthesis are from our previous implementation.

Table 3: Comparing Strategy 2 and Strategy 3 for literature models (s: seconds; d: days)

Models	ML degree X	degree $\mathcal{D}_{X,J}$	Sampling Timing		Total Timing	
			Strategy 2 (Algorithm 3)	Strategy 3 (Algorithm 6)	Strategy 2	Strategy 3
Model 1	3	6	0.05s	0.05s	1.7s	1.7s
Model 2	2	4	0.04s	0.05s	3.3s	3.9s
Model 3	4	14	0.05s	0.06s	370.3s	438.8s
Model 4	6	12	0.8s	0.3s	1501.0s	804.8s
Model 5	12	48	2.0s	4.8s	>1d	>2d
Model 6	10	34	62.1s	6.8s	>111178d	>13374d
Model 7	10	34	65.0s	6.9s	>144960d	>9203d
Model 8	23	102	26.8s	53.4s	>123d	>330d
Model 9	14	70	294.3s	122.2s	>7071130d	>2637947d

Table 3 compares Strategy 2 and Strategy 3 for Models 1–9. The column “Sampling Timing” gives the timings for doing sample once. The column “Total Timing” gives the timings for computing $\mathcal{D}_{X,J}$. The italics font timing means the computation did not finish in 2 hours, but we can estimate the sampling timing providing a lower bound (Example 7).

Example 7. The lower bound of the computational timings are approximated by the times in italics. For instance, consider Model 5 in Table 3. There are 5 parameters u_0, u_1, u_2, u_3, u_4 . By Algorithm 5, the total degree of $\mathcal{D}_{X,J}$ is 48 and the degrees with respect to the five parameters are: 31, 44, 48, 44 and 31. We check the timing for doing sample one time by Algorithm 3 is 2.0s. Then we estimate the timing for sampling of Strategy 2 is $2.0s \times 44 \times 31 \times 31 \approx 23.49h$ (1d) by formula (5) in Section 3.2.3. Similarly, we check the timing for doing sample one time by Algorithm 6 is 4.8s and estimate the total timing for sampling of Strategy 3 is $4.8s \times 44 \times 31 \times 31 \approx 56.37h$ (2d).

Remark 5. Whenever we compute an elimination ideal we use the Fgb command `fgb_gbasis_elim`. This is done in Algorithm 1-Line 1, Algorithm 3-Line 3, Algorithm 5-Lines 3, 7, Algorithm 6-Line 3, and Algorithm 7-Lines 3, 13.

5 Future Work and Final Application

5.1 Next steps

For our future work we want to determine the discriminants for larger sized problem such as Models 5–9. The interpolation method discussed in this paper is basically linear lifting interpolation.

Table 4: Largest intermediate matrix size after running algorithm 1 for 2 hours

Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
4213938×5121243	446567×635260	3193378×5474449	4314279×7141576	1990207×2328577	3835608×5561055

Section 3.2.3. Formula (5) gives the timing estimation for this linear lifting method. Although Strategy 3 improves T_s in formula (5), the interpolation is still expensive (see Table 3) since $d_2 \cdots d_n$ is quite a large number for these models. In order to overcome this problem, one possible way is to try Newton–Hensel lifting [14, 30].

We also want to develop heuristics for checking the output of probabilistic algorithm. For all examples we have tried in Section 4, the computational results of probabilistic algorithm are the same with the results computed by the standard algorithm as soon as both of algorithms finish the computation. For some examples such as Model 4, the standard algorithm gives no output in a reasonable time. However, there are ways to check the output of probabilistic algorithm. The first way to check is to simply run the algorithm a second time. The second way is to evaluate the discriminant on a random line. This univariate polynomial should agree by doing an elimination problem that replaces the data with the random line. The third way to check is to find a point on the discriminant. Solving the likelihood equations for this special point, we are able to check if solutions coincide by brute force. Moreover, if the discriminant has no rational points, then one could try to use numerical homotopy continuation methods instead.

The next step in this work is the complete positive root classification. The discriminant is necessary, but usually not enough for classifying positive solutions (see the next subsection regarding Model 4). In order to complete positive root classification, we have to consider the *discriminant sequence* [32].

5.2 Final Application

We end the paper with the discussion of real root classification on the 3×3 symmetric matrix model (Model 4).

Consider the following story with dice. A gambler has a coin, and two pairs of three-sided dice. The coin and the dice are all unfair. However, the two dice in the same pair have the same weight. He plays the same game 1000 rounds. In each round, he first tosses the coin. If the coin lands on side 1, he tosses the first pair of dice. If the coin lands on side 2, he tosses the second pair of dice. After the 1000 rounds, he records a 3×3 data matrix $[\bar{u}_{ij}]$ ($i, j = 1, 2, 3$) where \bar{u}_{ij} is the number of times for him to get the sides i and j with respect to the two dice. By the matrix $[\bar{u}_{ij}]$, he is trying to estimate the probability \bar{p}_{ij} of getting the sides i and j with respect to the two dice.

It is easy to check that the matrix

$$\begin{bmatrix} \bar{p}_{11} & \bar{p}_{12} & \bar{p}_{13} \\ \bar{p}_{21} & \bar{p}_{22} & \bar{p}_{23} \\ \bar{p}_{31} & \bar{p}_{32} & \bar{p}_{33} \end{bmatrix}$$

is symmetric and has at most rank 2. Let

$$p_{ij} = \begin{cases} \bar{p}_{ij} & i = j \\ \frac{1}{2}\bar{p}_{ij} & i < j \end{cases}, \quad u_{ij} = \begin{cases} \bar{u}_{ij} & i = j \\ \bar{u}_{ij} + \bar{u}_{ji} & i < j \end{cases}.$$

We have an algebraic statistical model $\mathcal{M} = \mathcal{V}(g) \cap \Delta_5$, whose Zariski closure is X where

$$g = \det \begin{bmatrix} 2p_{11} & p_{12} & p_{13} \\ p_{12} & 2p_{22} & p_{23} \\ p_{13} & p_{23} & 2p_{33} \end{bmatrix} \text{ and } \Delta_5 = \{(p_{11}, \dots, p_{33}) \in \mathbb{R}_{\geq 0}^6 \mid p_{11} + p_{12} + p_{13} + p_{22} + p_{23} + p_{33} = 1\}.$$

According to the Definition 2, we present the Langrange likelihood equations:

$$\begin{aligned}
F_0 &= p_{11}\lambda_1 + (8p_{22}p_{33} - 2p_{23}^2)p_{11}\lambda_2 - u_{11} \\
F_1 &= p_{12}\lambda_1 + (2p_{13}p_{23} - 4p_{12}p_{33})p_{12}\lambda_2 - u_{12} \\
F_2 &= p_{13}\lambda_1 + (2p_{12}p_{23} - 4p_{13}p_{22})p_{13}\lambda_2 - u_{13} \\
F_3 &= p_{22}\lambda_1 + (8p_{11}p_{33} - 2p_{13}^2)p_{22}\lambda_2 - u_{22} \\
F_4 &= p_{23}\lambda_1 + (2p_{12}p_{13} - 4p_{11}p_{23})p_{23}\lambda_2 - u_{23} \\
F_5 &= p_{33}\lambda_1 + (8p_{11}p_{22} - 2p_{12}^2)p_{33}\lambda_2 - u_{33} \\
F_6 &= g(p_{11}, p_{12}, p_{13}, p_{22}, p_{23}, p_{33}) \\
F_7 &= p_{11} + p_{12} + p_{13} + p_{22} + p_{23} + p_{33} - 1
\end{aligned}$$

where the unknowns are $p_{11}, p_{12}, p_{13}, p_{22}, p_{23}, p_{33}, \lambda_1, \lambda_2$ and parameters are $u_{11}, u_{12}, u_{13}, u_{22}, u_{23}, u_{33}$.

We have 8 equations in 8 unknowns with 6 parameters and the ML degree is 6 [19]. By the Algorithm 2, we have computed $\mathcal{D}_{X,J}$, which has 1307 terms with total degree 12. By a similar computation, we get $\mathcal{D}_{X,\infty}$ ⁴ whose factors are $g(u_{11}, \dots, u_{33})$ or positive when each u_i is positive.

For the data-discriminant \mathcal{D}_X we have computed above, we have also computed⁵ at least one rational point (sample point) from each open connected component of $\mathcal{D}_X \neq 0$ using **RAGlib**[28, 21, 15]. With these sample points, we can solve the real root classification problem on the open cells. By testing all 236 sample points, we see that if $g(u_{11}, \dots, u_{33}) \neq 0$, then

- if $\mathcal{D}_{X,J}(u_{11}, \dots, u_{33}) > 0$, then the system has 6 distinct real solutions and there can be 6 positive solution or 2 positive solutions;
- if $\mathcal{D}_{X,J}(u_{11}, \dots, u_{33}) < 0$, then the system has 2 distinct real (positive) solutions.

With 2 of these sample points, we see that the sign of \mathcal{D}_X is not enough to classify the positive solutions. For example, for the sample point $(u_{11} = 1, u_{12} = 1, u_{13} = \frac{280264116870825}{295147905179352825856}, u_{22} = 1, u_{23} = \frac{34089009205592922038535}{141080698675730650759168}, u_{33} = \frac{32898355113670387769001}{141080698675730650759168})$, the system has 6 distinct positive solutions. While for the sample point $(u_{11} = 1, u_{12} = 1, u_{13} = 199008, u_{22} = 30, u_{23} = 2022, u_{33} = 1)$, the system has also 6 real solutions but only 2 positive solutions.

6 Acknowledgments

We began this project in NIMS Thematic Program 2014 on Applied Algebraic Geometry. We had the idea for this extended version in 2015 summer before SIAM Conference: Applied Algebraic Geometry when the two authors were hosted by NIMS, Daejeon, Korea. We thank Professors Bernd Sturmfels, Hoon Hong, Jonathan Hauenstein and Frank Sottile for their valuable advice on this project. We thank Professors Mohab Safey EI Din and Jean-Charles Faugere for their software advice on **RAGlib** and **FGb** respectively. We thank the anonymous referees of ISSAC'15 for their insightful suggestions to greatly improve our work. We thank Professor Guillaume Moroz for his nice discussion and advice during ISSAC'15. We thank Professor Agnes Szanto for the reference on Shape Lemma (multivariate version).

References

- [1] N. Budur and B. Wang. The Signed Euler Characteristic of Very Affine Varieties. *Int. Math. Res. Not. IMRN*, (14):5710–5714, 2015.

⁴See \mathcal{D}_X on the website: <https://sites.google.com/site/rootclassification/publications/DD>

⁵The sample points were first successfully computed by one of the ISSAC'15 anonymous referees.

- [2] M.-L. G. Buot, S. Hoşten, and D. Richards. Counting and locating the solutions of polynomial systems of maximum likelihood equations, ii: The behrens-fisher problem. *Statistica Sinica*, 17:1343–1354, 2007.
- [3] F. Catanese, S. Hoşten, A. Khetan, and B. Sturmfels. The maximum likelihood degree. *Amer. J. Math.*, 128(3):671–697, 2006.
- [4] C. Chen, J. H. Davemport, J. P. May, M. M. Maza, B. Xia, and R. Xiao. Triangular decomposition of semi-algebraic systems. In *Proceedings of ISSAC’10*, pages 187–194. ACM New York, 2010.
- [5] G. E. Collins and H. Hong. *Partial Cylindrical Algebraic Decomposition for Quantifier Elimination*. Springer, 1998.
- [6] M. Coste and M. Shiota. Nash triviality in families of nash manifolds. *Inventiones Mathematicae*, 108(1):349–368, 1992.
- [7] D. A. Cox, J. Little, and D. Oshea. *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra*. Springer, 2007.
- [8] J. Draisma, E. Horobet, G. Ottaviani, B. Sturmfels, and R. R. Thomas. The Euclidean distance degree of an algebraic variety. *Found. Comp. Math.* 2014.
- [9] M. Drton, B. Sturmfels, and S. Sullivant. *Lectures on algebraic statistics*. Springer, 2009.
- [10] J.-C. Faugère. Fgb: A library for computing gröbner bases. In K. Fukuda, J. Hoeven, and M. Joswig, editors, *Mathematical Software - ICMS 2010*, volume 6327 of *Lecture Notes in Computer Science*, pages 84–87. Springer Berlin / Heidelberg, 2010.
- [11] J.-C. Faugère and S. Lachartre. Parallel gaussian elimination for gröbner bases computations in finite fields. In *Proceedings of the 4th International Workshop on Parallel and Symbolic Computation*, pages 89–97. ACM, 2010.
- [12] J.-C. Faugère, M. Safey EI Din, and P.-J. Spaenlehauer. Gröbner bases and critical points: the unmixed case. In *Proceedings of ISSAC’12*, pages 162–169. ACM New York, 2012.
- [13] M. Giusti, J. Heintz, J. E. Morais, J. Morgenstern, and L. M. Pardo. Straight-line programs in geometric elimination theory. *Journal of pure and applied algebra*, 124(1):101–146, 1998.
- [14] M. Giusti, G. Lecerf, and B. Salvy. A Gröbner free alternative for polynomial system solving. *Journal of Complexity*, 17:154–211, 2001.
- [15] A. Greuet and M. Safey EI Din. Probabilistic algorithm for the global optimization of a polynomial over a real algebraic set. *SIAM Journal on Optimization*, 24(3):1313–1343, 2014.
- [16] E. Gross, M. Drton, and S. Petrović. Maximum likelihood degree of variance component models. *Electronic Journal of Statistics*, 6:993–1016, 2012.
- [17] E. Gross and J. I. Rodriguez. Maximum likelihood geometry in the presence of data zeros. In *Proceedings of ISSAC’14*, pages 232–239. ACM New York, 2014.
- [18] J. Hauenstein, J. I. Rodriguez, and B. Sturmfels. Maximum likelihood for matrices with rank constraints. *Journal of Algebraic Statistics*, 5:18–38, 2014.
- [19] S. Hoşten, A. Khetan, and B. Sturmfels. Solving the likelihood equations. *Foundations of Computational Mathematics*, 5(4):389–407, 2005.

- [20] S. Hoşten and S. Sullivant. The algebraic complexity of maximum likelihood estimation for bivariate missing data. In *Algebraic and geometric methods in statistics*, pages 123–133. Cambridge University Press, 2009.
- [21] H. Hong and M. Safey EI Din. Variant quantifier elimination. *Journal of Symbolic Computation*, 47(7):883–901, 2012.
- [22] J. Huh. The maximum likelihood degree of a very affine variety. *Compos. Math.*, 149(8):1245–1266, 2013.
- [23] J. Huh and B. Sturmfels. *Likelihood geometry*, pages 63–117. Springer International Publishing, 2014.
- [24] Z. Jelonek. Testing sets for properness of polynomial mappings. *Mathematische Annalen*, 315:1–35, 1999.
- [25] D. Lazard and F. Rouillier. Solving parametric polynomial systems. *Journal of Symbolic Computation*, 42(6):636–667, 2005.
- [26] J. I. Rodriguez. Maximum likelihood for dual varieties. In *Proceedings of the 2014 Symposium on Symbolic-Numeric Computation*, SNC '14, pages 43–49, New York, NY, USA, 2014. ACM.
- [27] J. I. Rodriguez and X. Tang. Data-discriminants of likelihood equations. In *Proceedings of ISSAC'15*, pages 307–314. ACM, 2015.
- [28] M. Safey EI Din and E. Schost. Polar varieties and computation of one point in each connected component of a smooth real algebraic set. In *Proceedings of ISSAC'03*, pages 224–231. ACM Press, 2003.
- [29] M. Safey EI Din and E. Schost. Properness defects of projections and computation of in each connected component of a real algebraic set. *Discrete and Computational Geometry*, 32(3):417–430, 2004.
- [30] E. Schost. Computing parametric geometric resolutions. *Applicable Algebra in Engineering, Communication and Computing*, 13(5):349–393, 2003.
- [31] C. Uhler. Geometry of maximum likelihood estimation in gaussian graphical models. *Annals of Statistics*, 40(1):238–261, 2012.
- [32] L. Yang. Recent advances on determining the number of real roots of parametric polynomials. *Journal of Symbolic Computation*, 28(1):225–242, 1999.
- [33] L. Yang, X. Hou, and B. Xia. A complete algorithm for automated discovering of a class of inequality-type theorems. *Science in China Series F Information Sciences*, 44(1):33–49, 2001.

Appendix: Literature Models

Model 1. [9, Random Censoring Model]

$$2p_0p_1p_2 + p_1^2p_2 + p_1p_2^2 - p_0^2p_{12} + p_1p_2p_{12} = 0, \quad p_0 + p_1 + p_2 + p_{12} = 1$$

Model 2. [17, 3×3 Zero-Diagonal Matrix]

$$\det \begin{bmatrix} 0 & p_{12} & p_{13} \\ p_{21} & 0 & p_{23} \\ p_{31} & p_{32} & 0 \end{bmatrix} = 0, \quad p_{12} + p_{13} + p_{21} + p_{23} + p_{31} + p_{32} = 1$$

Model 3. [19, 17, Grassmannian of 2-planes in \mathbb{C}^4]

$$p_{12}p_{34} - p_{13}p_{24} + p_{14}p_{23} = 0, \quad p_{12} + p_{13} + p_{14} + p_{23} + p_{24} + p_{34} = 1$$

Model 4. [19, 3×3 Symmetric Matrix]

$$\det \begin{bmatrix} 2p_{11} & p_{12} & p_{13} \\ p_{12} & 2p_{22} & p_{23} \\ p_{13} & p_{23} & 2p_{33} \end{bmatrix} = 0, \quad p_{11} + p_{12} + p_{13} + p_{22} + p_{23} + p_{33} = 1$$

Model 5. [19, Bernoulli 3×3 Coin]

$$\det \begin{bmatrix} 12p_0 & 3p_1 & 2p_2 \\ 3p_1 & 2p_2 & 3p_3 \\ 2p_2 & 3p_3 & 12p_4 \end{bmatrix} = 0, \quad p_0 + p_1 + p_2 + p_3 + p_4 = 1$$

Model 6. [19, 3×3 Matrix]

$$\det \begin{bmatrix} p_{00} & p_{01} & p_{02} \\ p_{10} & p_{11} & p_{12} \\ p_{20} & p_{21} & p_{22} \end{bmatrix} = 0, \quad p_{00} + p_{01} + p_{02} + p_{10} + p_{11} + p_{12} + p_{20} + p_{21} + p_{22} = 1$$

Model 7. [17, Projection of 3×4 Matrix]

$$p_{12}p_{23}p_{34} - p_{12}p_{24}p_{33} - p_{13}p_{22}p_{34} + p_{13}p_{24}p_{32} + p_{14}p_{22}p_{33} - p_{14}p_{23}p_{32} = 0, \\ p_{12} + p_{13} + p_{14} + p_{22} + p_{23} + p_{24} + p_{32} + p_{33} + p_{34} = 1$$

Model 8. [19, Juke-Cantor Model, Example 18]

$$q_{000}q_{111}^2 - q_{011}q_{101}q_{110} = 0, \quad p_{123} + p_{dis} + p_{12} + p_{13} + p_{23} = 1$$

where

$$q_{111} = p_{123} + \frac{p_{dis}}{3} - \frac{p_{12}}{3} - \frac{p_{13}}{3} - \frac{p_{23}}{3}, \quad q_{110} = p_{123} - \frac{p_{dis}}{3} + p_{12} - \frac{p_{13}}{3} - \frac{p_{23}}{3}, \\ q_{101} = p_{123} - \frac{p_{dis}}{3} - \frac{p_{12}}{3} + p_{13} - \frac{p_{23}}{3}, \quad q_{011} = p_{123} - \frac{p_{dis}}{3} - \frac{p_{12}}{3} - \frac{p_{13}}{3} + p_{23}, \\ q_{000} = p_{123} + p_{dis} + p_{12} + p_{13} + p_{23}.$$

Model 9. [19, Example 16]

$$q_2q_7 - q_1q_8 = 0, \quad q_3q_6 - q_5q_4 = 0, \quad p_1 + p_2 + p_3 + p_4 + p_5 + p_6 + p_7 + p_8 = 1$$

where

$$q_1 = p_1 + p_2 + p_3 + p_4 + p_5 + p_6 + p_7 + p_8, \quad q_2 = p_1 - p_2 + p_3 - p_4 + p_5 - p_6 + p_7 - p_8, \\ q_3 = p_1 + p_2 - p_3 - p_4 + p_5 + p_6 - p_7 - p_8, \quad q_4 = p_1 - p_2 - p_3 + p_4 + p_5 - p_6 - p_7 + p_8, \\ q_5 = p_1 + p_2 + p_3 + p_4 - p_5 - p_6 - p_7 - p_8, \quad q_6 = p_1 - p_2 + p_3 - p_4 - p_5 + p_6 - p_7 + p_8, \\ q_7 = p_1 + p_2 - p_3 - p_4 - p_5 - p_6 + p_7 + p_8, \quad q_8 = p_1 - p_2 - p_3 + p_4 - p_5 + p_6 + p_7 - p_8.$$